Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220393

A Lightweight and Interpretable Model to Classify Bundle Branch Blocks from ECG Signals

Hugues TURBÉ[a,1], Mina BJELOGRLIC[a], Mehdi NAMDAR[b], Christophe GAUDET-BLAVIGNAC[a], Jamil ZAGHIR[a], Jean-Philippe GOLDMAN[a], Belinda LOKAJ[a] and Christian LOVIS [a]
[a] *Division of medical information sciences, University hospitals of Geneva and Department of radiology and medical informatics, University of Geneva, Switzerland*
[b] *Cardiology Division, University Hospitals of Geneva, Switzerland*

Abstract. Automatic classification of ECG signals has been a longtime research area with large progress having been made recently. However these advances have been achieved with increasingly complex models at the expense of model's interpretability. In this research, a new model based on multivariate autoregressive model (MAR) coefficients combined with a tree-based model to classify bundle branch blocks is proposed. The advantage of the presented approach is to build a lightweight model which combined with post-hoc interpretability can bring new insights into important cross-lead dependencies which are indicative of the diseases of interest.

Keywords. ECG automatic classification, Interpretability, Lightweight Model

1. Introduction

An electrocardiogram is a very common clinical exam, low-cost, non-invasive and allow to diagnose a range of cardiovascular diseases. Electrocardiograms measure the electric activity of the heart and aim to build a map of the heart in three orthogonal directions. Standard ECGs produce a signal with 12 derivations measured over approximately 10 seconds with a standard sampling frequency of 500Hz resulting in series with 5000 timesteps. ECGs are often analysed by non-experts staff and reports suggest that they are open to misdiagnosis [1]. This data abundance and a need for accurate diagnostic created a longtime interest for the automatic classification of diseases from electrocardiogram signals. Classification of time-series has often been seen as a difficult task for neural networks which struggle to learn long-term dependencies often encountered in ECG data. Recently, numerous models, mainly based on deep-neural networks, have achieved very high classification accuracy for a range of cardiovascular diseases [2,3].

One drawback of these approaches based on deep neural network is the lack of interpretability of the developed models. In this work, interpretability refers to building an understanding of which features were used and their relative importance for the model to classify a given ECG. Indeed with ever more complex models, it can be very difficult

[1] Corresponding Author: Hugues Turbé; E-mail: hugues.turbe@unige.

to understand how a model came to a given conclusion on a specific sample. Different methods have been developed to address this issue and provide what has been termed post-hoc explainability [4]. Post-hoc explainability is a key requirement for clinicians to adopt these technologies [5], but might also become a legal requirement following a recent recommendation from the European Commission [6]. Recent works looked at providing interpretability in the context of ECG classification using deep neural networks (DNN). The current approach consist in first training a DNN and then applying post-hoc explainability, principally LIME [7], to highlight the most important part of the signal used by the network to provide the classification of a given sample. The LIME interpretability method is widely used as it is model-agnostic, meaning it can be applied on any architecture. LIME and other interpretability methods were initially developed for natural language processing (NLP) or image classification task. Given this background, one current limitation is that they fail to highlight important dependencies captured by the network both across time steps and different leads of the ECG. With this issue in mind, the presented method aims to address the following issues:

- The computing intensive nature of deep neural network required to learn the complex dependencies observed in ECG signals.
- The lack of interpretability method to highlight dependencies across leads captured by the models.

As part of this research, it was decided to build models to classify the presence of left and right bundle branch blocks respectively LBBB and RBBB. Bundle branch blocks refer to delay or blockage of the electrical signal responsible for making the heart beat [8]. Bundle branch blocks are associated with higher risk of serious cardiovascular complications and mortality in specific conditions [9,10]. While automatic classification of bundle branch blocks is well studied [11], the following work aims to present a much smaller model, which does not rely on CNN or LSTM layers, with improved interpretability of the classification models.

2. Methods

The CPSC subset from the large *PhysioNet: Classification of 12-lead ECGs* dataset [12] was used. This dataset includes a total of 6850 ECGs, with a minimal length of 10 seconds, annotated for 9 cardiovascular diseases. Out of these ECGs, 232 were annotated with a LBBB diagnostic, 1854 with a RBBB and 916 with a normal rhythm.

The model presented in this work is based on using a multivariate autoregressive model (MAR) as a preprocessing step to capture inter-lead dependencies. MAR model aims to explain current values of a multivariate time series as a linear combination of the past values. A MAR model of order p will use the p previous time step of the serie of interest to predict the following. For a time series $\boldsymbol{X} \in \mathbb{R}^{M \times T}$ where M is the number of feature of the time series, in our case the number of derivation obtained with the ECG signals, and T the number of time steps in the series, we can write our MAR model as follows:

$$\boldsymbol{X_t} = \sum_{l=1}^{p} \boldsymbol{A_i} \cdot \boldsymbol{X_{t-i}} + \epsilon_t \quad (1)$$

Where $\boldsymbol{A_i} \in \mathbb{R}^{M\times M}$ is the matrix of coefficient and ϵ_t the additive Gaussian noise with zero mean and covariance R. For the rest of the analysis, a MAR model of order one is used, so that $\boldsymbol{A} = \boldsymbol{A_i}$. The matrix $\boldsymbol{A}$ is central to the presented analysis as it encodes the conditional dependencies across the leads of the ECG signals [13] as well as reducing the dimension from the domain $\mathbb{R}^{5000\times12}$ to a smaller $\mathbb{R}^{12\times12}$ domain. This dimensionality reduction allows to greatly reduce the computing power required to train a classification model. The rest of the analysis relies on the assumption that the changes induced by a left or right bundle branch will affect the dependency across the leads and hence can be used as a marker to classify these two diseases. MAR models have been widely adopted to infer organisation model of the human brain using functional magnetic resonance imaging (fMRI) time series [14]. Surprisingly these models have not yet been used, to the authors' knowledge, on ECG data. As a prerequisite to compute the MAR coefficients, the stationarity of the time series was tested using the Augmented Dickey–Fuller test. In addition, all samples were normalised by removing their mean and scaling them to unit variance.

A first statistical analysis of the significance of the MAR model's coefficients with regards to the diseases of interest was performed using a permutation test based on the means of the groups [15].

For each of the disease of interest, the MAR coefficients were used as input to train a tree-based model, more specifically a Light Gradient Boosting Model (LightGBM) [16]. LightGBM is based on gradient boosted tree model and has gained attraction for its state-of-the-art performance. For each of the two trained models the target for each sample was whether the latter was annotated with the disease of interest or as a sample with a sinus rhythm. Once these models are trained, *Tree SHAP* [17] is used to provide post-hoc interpretability under the form of Shapley values. Shapley values aim to explain which coefficients of the MAR model were used by the trained tree model as indicative of the diseases of interest.[2]

3. Results

Initial permutation tests showed that out of the 144 (12×12) MAR coefficients, 66 for LBBB and 84 for RBBB had a significant distribution difference with regards to samples with a sinus rhythm $p \leq 0.05$. Hyperparameter optimisation was performed for the two classification tasks with the optimal hyperparameters for each model along resulting classification metrics presented respectively in Table 1 and Table 2.

Table 1. Model's Hyperparameters

	LBBB	RBBB
Learning rate	0.12	0.11
Nb. leaves	1420	1220
Max depth	4	6

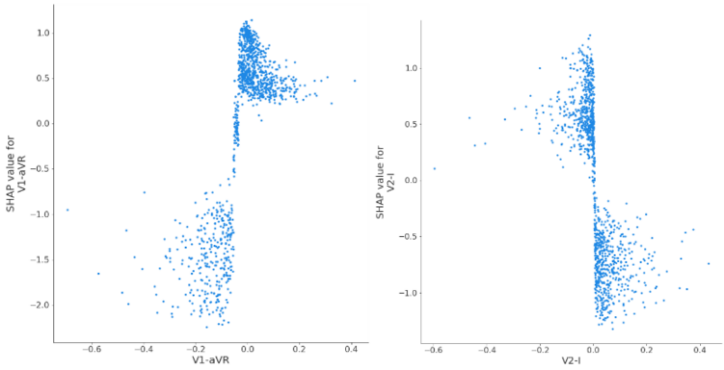
Table 2. Classification Metrics

	LBBB	RBBB
Accuracy	0.97	0.84
Precision	0.95	0.85
Recall	0.88	0.91
F1	0.91	0.88

Shapley values were calculated using TreeShap [17] for the two models. Shapley values explain how each feature across the dataset influenced the prediction of the model of

[2] Full implementation of the discussed method is presented at https://github.com/hturbe/ECG_MAR_Model

interest. In Figure 1, the influence of two important MAR coefficients on the classification of LBBB are shown. This figure shows the Shapley value attributed to a specific feature across the entire dataset and allows to gain insights into how specific values of the feature influence the model's predictions. The magnitude of the Shapley value shown on the y-axis reflects the importance of this feature for the model to classify the sample. In addition, a positive Shapley value indicates that this feature played in favour of the sample being classified as having the disease of interest while a negative Shapley value indicates that the feature played against the sample being predicted with the disease. Figure 1a shows that a positive dependence between lead V1 and aVR is associated by the network as an indication of a LBBB while the opposite is true for the dependence between leads V2 and I as shown by Figure 1b. Regarding RBBB, the dependence between leads V1 and V2 was found to be a key diagnostic criterion.



(A) (B)

Figure 1. Shapley values computed on the model trained to predict LBBB. Shapley values are presented for the MAR coefficient encoding the dependence between leads V1-aVR (a) and leads V2-I (b).

4. Conclusions

The presented approach is based on encoding the temporal dependencies across the ECG leads within the MAR model. This initial step allows to reduce each ECG into a matrix of 144 coefficients. The coefficients are then used as inputs for tree-based models with smaller computing requirements than traditional approaches based on neural networks architectures. Two models are trained achieving an accuracy of respectively 97% for the classification of LBBB and 84% for RBBB. Of interest, the accuracy for the classification of LBBB is achieved with a dataset including only 232 samples annotated with the disease. One current drawback of the method is that it discards "intra-lead" feature, such as QRS length and others. An avenue to fix this issue would be to augment the MAR coefficients with intra-signal measurements.

Shapley values provide interesting insights into how a change in the dependence between a pair of leads are indicative of a specific disease. The research showed how change in dependencies between lead V1 and aVR and leads V2 and I were discriminative to predict a LBBB. In addition, while some research have focused on developing interpretability methods for ECG classification, they show time steps in given leads which were important for the classification. This information is interesting but does not reflect how dependence across leads, which might be captured by the developed models, play a role in the final prediction.

Finally, the models presented in this article should be trained and tested on a more extensive dataset. Indeed, currently the classification is limited between sinus rhythm and the two diseases which were analysed. However this classification does not reflect a clinical setting where patient might also exhibit different cardiovascular diseases.

References

[1] Awan M, Omar M, Omar BA, Malozzi C, Awan GM. Abstract 244: ECG Recording Errors Prevalence in a Hospital Setting, *Circulation: Cardiovascular Quality and Outcomes*. **11** (2018) A244–A244. doi:10.1161/circoutcomes.11.suppl_1.244.

[2] Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network, *Nat Commun*. **11** (2020) 1760. doi:10/drkd.

[3] Liang Y, Yin S, Tang Q, Zheng Z, Elgendi M, Chen Z. Deep Learning Algorithm Classifies Heartbeat Events Based on Electrocardiogram Signals, *Front. Physiol.* **11** (2020) 569050. doi:10/gm3srb.

[4] Lipton ZC.The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery., *Queue*. **16** (2018) 31–57. doi:10/gkbfrd.

[5] Shad R, Cunningham JP, Ashley EA, Langlotz CP, Hiesinger W. Designing clinically translatable artificial intelligence systems for high-dimensional medical imaging, *Nat Mach Intell*. **3** (2021) 929–935. doi:10/gnn5hz.

[6] European Commission, Directorate-General for Communications Networks, Content and Technology, Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts COM/2021/206 final, (2021).

[7] Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier, *ArXiv:1602.04938 [Cs, Stat]*. (2016). http://arxiv.org/abs/1602.04938 (accessed March 2, 2021).

[8] McAnulty JH, Rahimtoola SH. Bundle branch block, *Progress in Cardiovascular Diseases*. **26** (1984) 333–354. doi:10.1016/0033-0620(84)90009-4.

[9] Tan NY, Witt CM, Oh JK, Cha YM. Left Bundle Branch Block: Current and Future Perspectives, *Circ: Arrhythmia and Electrophysiology*. **13** (2020). doi:10.1161/CIRCEP.119.008239.

[10] Bussink BE, Holst AG, Jespersen L, Deckers JW, Jensen GB, Prescott E. Right bundle branch block: prevalence, risk factors, and outcome in the general population: results from the Copenhagen City Heart Study, *European Heart Journal*. **34** (2013) 138–146. doi:10.1093/eurheartj/ehs291.

[11] Oh SL, Tan EYKNgRS, Acharya UR. Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats, *Computers in Biology and Medicine*. **102** (2018) 278–287. doi:10/gfnqsp.

[12] Perez Alday EA, Gu A, Shah A, Robichaux C, Wong AKI, Liu C, et al. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020, Cardiovascular Medicine, 2020. doi:10.1101/2020.08.11.20172601.

[13] Lütkepohl H, New introduction to multiple time series analysis, Springer Science & Business Media, 2005.

[14] Rogers BP, Katwal SB, Morgan VL, Asplund CL, Gore JC. Functional MRI and multivariate autoregressive models, *Magnetic Resonance Imaging*. **28** (2010) 1058–1065.

[15] Wilcox RR, 8 - COMPARING TWO INDEPENDENT GROUPS, in: R.R. Wilcox (Ed.), Applying Contemporary Statistical Techniques, Academic Press, Burlington, 2003: pp. 237–284. doi:10.1016/B978-012751541-0/50029-8.

[16] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc., 2017. https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

[17] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees, *Nature Machine Intelligence*. **2** (2020) 56–67.